

Rates of Convergence to Stationarity for Multidimensional RBM

Blanchet, J. and Chen, X.

August 15, 2016

Abstract

We provide the first rate of convergence analysis for reflected Brownian motion (RBM) as the dimension grows under natural uniformity conditions. In particular, if the underlying routing matrix is uniformly contractive, uniform stability of the drift vector holds, and the variances of the underlying Brownian Motion (BM) are bounded, then we show that the RBM converges exponentially fast to stationarity with a relaxation time of order $O(d^4(\log(d))^2)$ as the dimension $d \rightarrow \infty$.

1 Introduction

Multidimensional Reflected Brownian Motion (RBM) was introduced in [4] and it is one of the most important models in Operations Research because it can be used to approximate (in distribution) the workload content of a very large class of stochastic networks of interest as the traffic utilization of the system approaches 100% (i.e. in heavy traffic). See Chapter 7 of [2] and the references therein. Moreover, it has been shown that the approximation holds also for the underlying steady-state distributions in significant generality (see [1] and [3]).

In this paper, we study the rate of convergence to stationarity of multidimensional RBM. We provide the first rate of convergence analysis for RBM as the dimension d grows under natural uniformity conditions. In particular, if the underlying routing matrix is uniformly contractive (see Assumption A1), uniform stability of the drift vector holds (see Assumption A2), and the variances of the underlying Brownian Motion (BM) are bounded (see Assumption A3), then we show that the RBM converges exponentially fast to stationarity with a relaxation time of order $O(d^4(\log(d))^2)$ as $d \rightarrow \infty$.

In Section 2, we first introduce our notation and provide the statement of our main result. Also in Section 2, we provide a step-by-step strategy behind the proof of our main result. The proof is divided into three steps, which are developed throughout Sections 3 to 5.

2 Notation, Assumptions and Main Result

We start this section by explaining the motivation and definition of RBM and the assumptions that we shall impose throughout the paper. We concentrate on the case where $d \geq 2$, and the case in which $d = 1$ is standard.

2.1 Notation

For convenience, we summarize the common notations used through out the paper. We shall use boldface to write vector quantities, which are encoded as columns. For instance, we write $\mathbf{y} = (y_1, \dots, y_d)^T$. We use $\mathbf{1}$ to denote the vector with all entries equal to unity. We define the following norms of vectors: $\|\mathbf{y}\|_\infty = \max_{i=1}^d |y_i|$ and $\|\mathbf{y}\|_1 = \sum_{i=1}^d |y_i|$.

We write I to denote the identity matrix. For a $d \times d$ matrix A , we let A^T be its transposition. For any subsets S_1 and S_2 of $\{1, 2, \dots, d\}$, we write $A_{S_1 S_2}$ as the submatrix of A such that $A_{S_1 S_2} = \{A_{ij} : i \in S_1, j \in S_2\}$. Similarly, $\mathbf{y}_{S_1} = (y_i : i \in S_1)$ and $A_{S_1} = \{A_{ij} : i \in S_1, 1 \leq j \leq d\}$.

All inequalities involving vectors or matrices are understood componentwise. For example, $\mathbf{y} \geq \mathbf{z}$ means that $y_i \geq z_i$ for all $i \in \{1, 2, \dots, d\}$.

For any subset S of $\{1, 2, \dots, d\}$, \bar{S} represents its compliment set, i.e., $\bar{S} = \{1 \leq i \leq d : i \notin S\}$. For all $1 \leq i, j \leq d$, δ_{ij} is the Kronecker delta, i.e., $\delta_{ij} = 1$ if $i = j$, and $\delta_{ij} = 0$ if $i \neq j$. The arrow “ \implies ” represents convergence in distribution. The equality $A \stackrel{D}{=} B$ means that A and B are equal in distribution. We use $N(0, 1)$ to refer to a generic standard normal random variable.

2.2 Motivation, Definition of RBM, and Assumptions

Let us consider the stochastic fluid network model introduced by [6]. It is a network of d queueing stations indexed by $\{1, 2, \dots, d\}$. Jobs arrive to the network according to some counting process $(N(t) : t \geq 0)$. The k -th arrival brings a vector of job requirements $\mathbf{W}(k) = (W_1(k), \dots, W_d(k))^T$, which adds $W_i(k)$ units of workload to the i -th station right at the moment of arrival, for $i \in$

$\{1, \dots, d\}$.

From the previous description, we know that the total amount of work that arrives to the i -th station, up to and including time t , is denoted by

$$J_i(t) = \sum_{k=1}^{N(t)} W_i(k). \quad (1)$$

Let us now assume that for all $i \in \{1, \dots, d\}$, the server of station i processes the workload as a fluid at rate $r_i > 0$. That means, if the workload in the i -th station remains strictly positive during the time interval $[t, t+h]$, the output from station i during this time interval will be $r_i h$. In addition, for all $1 \leq i, j \leq d$, let $Q_{i,j} \geq 0$ be the proportion of the fluid circulated to the j -th station, after being processed by the i -th server. The matrix $Q = (Q_{i,j} : 1 \leq i, j \leq d)$ is called the routing matrix of the network. Without loss of generality, we assume that $Q_{i,i} = 0$. We introduce an extra notation $Q_{i,0} = 1 - \sum_{j=1}^d Q_{i,j} \geq 0$ to represent the proportion of the fluid that leaves the network immediately after being processed by the i -th sever. Note that the matrix Q does not include $Q_{i,0}$.

It is natural to assume that arriving jobs will eventually leave the network, which is equivalent to assuming that $Q^n \rightarrow 0$ as $n \rightarrow \infty$; which, in turn, is equivalent to requiring that Q be a strict contraction in the sense that it has a spectral radius which is strictly less than one. In other words, one assumes there exists $\beta \in (0, 1)$ and $\kappa \in (0, \infty)$ such that:

$$\|\mathbf{1}^T Q^n\|_\infty \leq \kappa (1 - \beta)^n. \quad (2)$$

The dynamics of such a stochastic fluid network can be expressed formally in differential notation as follows. Let $Y_i(t)$ denote the workload content of the i -th station at time t , then given $Y_i(0)$, we write:

$$\begin{aligned} dY_i(t) &= dJ_i(t) - r_i I(Y_i(t) > 0) dt + \sum_{j:j \neq i} Q_{j,i} r_j I(Y_j(t) > 0) dt \\ &= dJ_i(t) - r_i dt + \sum_{j:j \neq i} Q_{j,i} r_j dt + r_i I(Y_i(t) = 0) dt \\ &\quad - \sum_{j:j \neq i} Q_{j,i} r_j I(Y_j(t) = 0) dt \end{aligned} \quad (3)$$

for $i \in \{1, \dots, d\}$. These equations take a neat form in matrix notation. Let $\mathbf{r} = (r_1, \dots, r_d)^T$ be the column vector corresponding to the service rates, and define the so-called “reflection matrix” as $R = (I - Q)^T$. Let

$$\mathbf{X}(t) = \mathbf{J}(t) - R\mathbf{r}t,$$

where $\mathbf{J}(t)$ is a column vector with its i -th component equal to $J_i(t)$ as defined in (1), then we can see from (3) that $\mathbf{Y}(\cdot)$ solves the following stochastic differential equation (SDE) with constraints known as the Skorokhod problem.

Skorokhod Problem: *Given a process $\mathbf{X}(\cdot)$ and a matrix R , we say that the pair (\mathbf{Y}, \mathbf{L}) solves the associated Skorokhod problem if*

$$0 \leq \mathbf{Y}(t) = \mathbf{Y}(0) + \mathbf{X}(t) + R\mathbf{L}(t), \quad \mathbf{L}(0) = 0 \quad (4)$$

where the i -th entry of $\mathbf{L}(\cdot)$ is non-decreasing and $\int_0^t Y_i(s) dL_i(s) = 0$.

The seminal paper [4] shows that the Skorokhod problem has a unique solution when the input $\mathbf{X}(\cdot)$ is continuous and R is a so-called M -matrix. In particular, a matrix R is said to be an M -matrix if

$$R^{-1} \text{ exists and it has non-negative entries.} \quad (5)$$

In our case, $\mathbf{X}(\cdot)$ is a multi-dimension Brownian motion with drift vector $\boldsymbol{\mu}$ and covariance matrix $\Sigma := CC^T$, and hence it is continuous almost surely. The reflection matrix $R = (I - Q)^T$ is indeed an M -matrix. The unique solution to the Skorokhod problem when the input is a $(\boldsymbol{\mu}, \Sigma)$ -Brownian Motion is called a $(\boldsymbol{\mu}, \Sigma, R)$ -RBM.

To understand intuitively why the M -condition assumption is very natural, once again we go back to the stochastic fluid network depicted in (3) and note that $R = I - Q^T$ being an M -matrix is equivalent to requiring that (2) holds.

To appreciate the delicate nature of $\mathbf{L}(\cdot)$, note that in the setting of the stochastic fluid network depicted in (3) we have that

$$L_i(t) = \int_0^t r_i I(Y_i(s) = 0) ds. \quad (6)$$

For general Skorokhod problems, under the M -condition and some mild conditions on $\mathbf{X}(\cdot)$,

the assumption that

$$R^{-1}E\mathbf{X}(1) < 0, \quad (7)$$

implies that $\mathbf{Y}(t) \implies \mathbf{Y}(\infty)$ as $t \rightarrow \infty$, where $\mathbf{Y}(\infty)$ is a random variable with the (unique) stationary distribution of $\mathbf{Y}(\cdot)$. In particular, according to [5], condition (7) is necessary and sufficient for stability of the $(\boldsymbol{\mu}, \Sigma, R)$ -RBM (i.e. a unique stationary distribution exists) under the M -condition (5).

In this paper, we shall consider a family of $(\boldsymbol{\mu}, \Sigma, R)$ -RBMs indexed by the dimension d . Implicitly, then, R , $\boldsymbol{\mu}$, and Σ are indexed by their dimension. Our goal is to derive rates of convergence to stationarity that behave gracefully as $d \rightarrow \infty$ under suitable uniformity conditions, which are stated in the following assumptions.

Assumptions:

A1) Uniform contraction: We let $R = I - Q^T$, where Q is substochastic and assume that there exists $\beta_0 \in (0, 1)$ and $\kappa_0 \in (0, \infty)$ independent of d such that

$$\|\mathbf{1}^T Q^n\|_\infty \leq \kappa_0(1 - \beta_0)^n. \quad (8)$$

Under (8) we observe that

$$\|R^{-1}\mathbf{1}\|_\infty \leq b_1 := \kappa_0/\beta_0 < \infty.$$

A2) Uniform stability: We write $\mathbf{X}(t) = \boldsymbol{\mu}t + C\mathbf{B}(t)$, where $\mathbf{B}(t) = (B_1(t), \dots, B_d(t))^T$ and the $B_i(\cdot)$'s are standard Brownian motions, and the matrix C satisfies $\Sigma = CC^T$. We assume that there exists $\delta_0 > 0$ independent of d such that

$$R^{-1}\boldsymbol{\mu} < -\delta_0\mathbf{1}.$$

A3) Uniform marginal variability: Define $\sigma_i^2 = \Sigma_{i,i}$ (i.e. the variance of the i -th coordinate of \mathbf{X}). We assume that there exists $b_0 \in (0, \infty)$, independent of $d \geq 1$, such that

$$b_0^{-1} \leq \sigma_i^2 \leq b_0.$$

Remark: An important constant to be used in the sequel is $\delta_1 = \delta_0 \beta_0 / (2\kappa_0)$. This constant will be used in the introduction of a useful dominating process.

We recognize that there are many ways in which one can embed a family of RBM's increasing in dimensionality. Our assumptions, we believe, constitute a reasonable departing point to rates of convergence to stationarity for large networks. Under condition (7), as mentioned earlier, there is a unique stationary distribution for the process \mathbf{Y} . Assumptions A1) and A2) are natural uniform extensions of (2) and (7). Assumption A3), we believe, is also natural. The lower bound in A3) simply avoids degeneracies. The upper bound can be seen as an assumption of tightness of the marginal steady-state distributions. If one believes that any given node in the network can be approximated by a general single-server queue in heavy traffic, then Assumption A3) would guarantee that the steady-state distributions of those nodes in isolation remain tight uniformly in d .

2.3 The Main Result: Statement

In order to quantify the rate of convergence to stationarity of RBM, we shall use Wasserstein's distance. Let us define

$$\mathcal{L} = \{f : R^d \rightarrow R \text{ such that } |f(x) - f(y)| \leq \|x - y\|_\infty\}.$$

In other words, \mathcal{L} is the set of Lipschitz continuous functions on R^d with the Lipschitz constant equal to one under the uniform norm. Suppose that the random variable $\mathbf{U} \in R^d$ has distribution ν in R^d and that $\mathbf{V} \in R^d$ has distribution ϖ . The associated Wasserstein distance (of order 1) between ν and ϖ is defined as

$$d_W(\nu, \varpi) = \sup_{f \in \mathcal{L}} |Ef(\mathbf{U}) - Ef(\mathbf{V})|.$$

With a slight abuse of notation, we shall actually write $d_W(\mathbf{U}, \mathbf{V})$ instead of $d_W(\nu, \varpi)$. We have chosen the Wasserstein distance of order 1 because in the stochastic network setting (which provides some of the main applications motivating the use of RBM), Lipschitz continuous functions of the underlying process are natural quantities to study. Examples of these functions include the

maximum workload and the total workload in a subset of stations in the network. Our results, therefore, allow us to immediately quantify initial transient errors in expectations of this sort.

Our main result is the following:

Theorem 1. *Under assumptions A1) to A3), for any $\beta \in (0, \min(\beta_0, 1/3) \cdot 1/3)$ satisfying,*

$$P\left(N(0, 1) < \sqrt{b_0}(\delta_0 - b_1^2)\right) \geq \beta/d, \quad (9)$$

we have that

$$\begin{aligned} d_W(\mathbf{Y}(t), \mathbf{Y}(\infty)) &\leq 3 \cdot d \cdot \exp\left(-\zeta_1 \cdot \frac{t}{(d^4 \log(d))}\right) \\ &\quad \cdot \left(\kappa_0 \cdot \|\mathbf{Y}\|_1 \cdot \exp\left(\zeta_0 \cdot \frac{\|\mathbf{Y}\|_\infty}{d^3 \log(d)}\right) + \frac{\kappa_0^{1/2}}{\delta_0^{1/2} \beta_0^{1/2}} b_0^{1/2}\right), \end{aligned} \quad (10)$$

as $t \rightarrow \infty$. Here ζ_0 and ζ_1 are two constants independent of d :

$$\zeta_0 = \frac{\delta_1 \cdot \beta^2}{2 \max_{i=1}^d \sigma_i^2}, \quad \zeta_1 = \frac{\delta_1^2 \cdot \beta^2}{16 \max_{i=1}^d \sigma_i^2}.$$

In particular, the relaxation time of RBM is of order $O\left(d^4 (\log(d))^2\right)$. (The relaxation time, $t^(d)$, satisfies*

$$d_W(\mathbf{Y}(t^*(d)), \mathbf{Y}(\infty)) \leq 1/2.)$$

Remark: We can actually relax Assumption A1) and allow the contraction bound b_1 to increase with d , as long as (9) holds. In particular, if we make $b_1 = O\left(\log(d)^{1/4}\right)$, then we can choose $\beta = O(d^{-\gamma})$ for some $\gamma > 0$ and we still obtain that the relaxation time $t^*(d)$ is polynomial in d (assuming that the rest of the assumptions remain in place). It appears that the contraction bound b_1 has the most impact on the speed of convergence to stationarity.

2.4 The Main Result: Strategy of the Proof

We first explain the main steps in the proof of Theorem 1. All the details, including the technical lemmas will be given in the following sections.

Step 0: We start by considering a natural coupling. Given the underlying $(\boldsymbol{\mu}, \Sigma)$ -Brownian

motion $\mathbf{X}(\cdot)$, we consider the $(\boldsymbol{\mu}, \Sigma, R)$ -RBM, $\mathbf{Y}(\cdot)$, obtained by solving the Skorokhod problem with reflection matrix R in (4). In order to emphasize the dependence on the initial condition, we will also write $\mathbf{Y}(t; \mathbf{Y}(0)) := \mathbf{Y}(t)$. Now let us use $\mathbf{Y}(\infty)$ to denote a random variable with the stationary distribution of $\mathbf{Y}(\cdot)$ but independent of $\mathbf{X}(\cdot)$. We then have, by stationarity, that

$$\mathbf{Y}(\infty) \stackrel{D}{=} \mathbf{Y}(t; \mathbf{Y}(\infty)).$$

We consider the process $\mathbf{Y}(\cdot; \mathbf{Y}(0))$ coupled with $\mathbf{Y}(\cdot; \mathbf{Y}(\infty))$, where the driving signal, $\mathbf{X}(\cdot)$, is common to both processes, but the initial conditions are different.

Note that for any $f \in \mathcal{L}$,

$$|Ef(\mathbf{Y}(t; \mathbf{Y}(0))) - Ef(\mathbf{Y}(t; \mathbf{Y}(\infty)))| \leq E \|\mathbf{Y}(t; \mathbf{Y}(0)) - \mathbf{Y}(t; \mathbf{Y}(\infty))\|_1$$

and hence

$$d_W(\mathbf{Y}(t; \mathbf{Y}(0)), \mathbf{Y}(t; \mathbf{Y}(\infty))) \leq E \|\mathbf{Y}(t; \mathbf{Y}(0)) - \mathbf{Y}(t; \mathbf{Y}(\infty))\|_1. \quad (11)$$

Therefore, to prove Theorem 1, it suffices to show that

$$E \|\mathbf{Y}(t; \mathbf{Y}(0)) - \mathbf{Y}(t; \mathbf{Y}(\infty))\|_1$$

can be bounded by the right hand side of (10). We shall do this through the following steps.

Step 1: The first step in the proof involves bounding

$$\|\mathbf{Y}(t; \mathbf{Y}(0)) - \mathbf{Y}(t; \mathbf{Y}(\infty))\|_1.$$

Define $\eta^0(\mathbf{y}) = 0$,

$$\begin{aligned} \eta_i^k(\mathbf{y}) &= \inf\{t > \eta^{k-1}(\mathbf{y}) + 1 : Y_i(t; \mathbf{y}) = 0\}, \\ \eta^k(\mathbf{y}) &= \sup\{\eta_i^k(\mathbf{y}) : 1 \leq i \leq d\}, \end{aligned} \quad (12)$$

and write

$$\mathcal{N}(t; \mathbf{y}) = \sup\{k \geq 0 : \eta^k(\mathbf{y}) \leq t\}.$$

We will show that

$$\begin{aligned}
& \|\mathbf{Y}(t; \mathbf{Y}(\infty)) - \mathbf{Y}(t; \mathbf{Y}(0))\|_1 \\
& \leq \|\mathbf{Y}(t; \mathbf{Y}(\infty)) - \mathbf{Y}(t; \mathbf{0})\|_1 + \|\mathbf{Y}(t; \mathbf{Y}(0)) - \mathbf{Y}(t; \mathbf{0})\|_1 \\
& \leq d \cdot \kappa_0 \cdot ((1 - \beta_0)^{\mathcal{N}(t; \mathbf{Y}(\infty))} \|\mathbf{Y}(\infty)\|_1 + (1 - \beta_0)^{\mathcal{N}(t; \mathbf{Y}(0))} \|\mathbf{Y}(0)\|_1).
\end{aligned} \tag{13}$$

(13) is obtained based on some elementary estimates following the analysis in [7]. Intuitively, we show that when all of the coordinates have hit zero at least once, the difference $\mathbf{Y}(t; \mathbf{Y}(\infty)) - \mathbf{Y}(t; \mathbf{Y}(0))$ shrinks by a factor which can be expressed in terms of a suitable product of substochastic matrices.

Step 2: Combining (11) and (13), it is easy to see that the key to our estimates involves bounding $E[(1 - \beta_0)^{\mathcal{N}(t; \mathbf{y})}]$ and $\|\mathbf{Y}(\infty)\|_1$.

At this point, we invoke a well-known sample-path upper bound $\mathbf{Y}^+(t; \mathbf{y})$ for $\mathbf{Y}(t; \mathbf{y})$ (see Lemma 3.1 in [6]). In particular, $\mathbf{Y}^+(\cdot; \mathbf{y})$ is also a RBM with its reflection matrix equal to the identity matrix, and it dominates $\mathbf{Y}(t; \mathbf{y})$ in the sense that $R^{-1}\mathbf{Y}^+(t; \mathbf{y}) \geq R^{-1}\mathbf{Y}(t; \mathbf{y})$ for all t . Besides, $\mathbf{Y}(\cdot; \mathbf{y})$ has a unique stationary distribution regardless of the initial condition \mathbf{y} . Let $\mathbf{Y}^+(\infty)$ follow the stationary distribution of $\mathbf{Y}^+(\cdot)$, then it is well-understood that $Y_i^+(\infty)$ follows an exponential distribution with mean $E[Y_i^+(\infty)] = \sigma_i^2/2(\mu_i^+ - \mu_i)$ marginally. Therefore, using Assumptions A1) - A3), one can show that $\sup_{i \geq 1} E[Y_i^+(\infty)] < \infty$. This upper bound process, together with Steps 1 and 2, already hints at the polynomial-time nature of the relaxation time. For example, if Σ is diagonal, a straightforward calculation shows that $E[\max_{1 \leq i \leq d} Y_i^+(\infty)] = O(\log(d))$. On the other hand, starting from equilibrium, in a time interval of order $O(d)$ the maximum coordinate fluctuates at most $O(\log(d))$ units, while, with very high probability, all coordinates will hit zero at least once during this time (due to the negative drift of the underlying Brownian motion driving \mathbf{Y}^+). One might expect that the coordinates of the lower bound process would also have visited zero during this time. However, such a reasoning is not implied by the type of domination that can be guaranteed between $\mathbf{Y}^+(t; \mathbf{y})$ and $\mathbf{Y}(t; \mathbf{y})$. In addition, the matrix Σ is not diagonal. So, due to all of these complications, the quantitative bounds become somewhat involved. The strategy to bound $E[(1 - \beta_0)^{\mathcal{N}(t; \mathbf{y})}]$ is split into several substeps.

Step 2.1 (estimating the time to visit a compact): First, we define $\tau^+(\mathbf{y}) = \inf\{t \geq 0 :$

$\mathbf{Y}^+(t; \mathbf{y}) \leq \mathbf{1}\}$. We define a suitable function $h(\mathbf{y}; \theta) \geq 0$ which behaves like $\theta \|\mathbf{y}\|_\infty$ for small θ . For each θ small enough, we can find $\chi(\theta) > 0$ such that

$$E \left[\exp \left(\chi(\theta) \tau^+(\mathbf{y}) \right) \right] \leq \exp \left(h(\mathbf{y}; \theta) \right),$$

and $h(\mathbf{y}; \theta) + \chi(\theta) \rightarrow 0$ as $\theta \rightarrow 0$. It turns out that $\chi(\theta) = O(\theta/d)$. Step 2.1 is executed by means of a suitable Lyapunov argument.

Step 2.2 (geometric trials for visits to zero): Step 2.1 allows us to estimate the time until all of the components of the process $\mathbf{Y}(\cdot)$ are inside a compact set (this is due to the domination property of \mathbf{Y}^+ and Assumption A2)). Then, using a geometric trial argument, we estimate the time it takes for the d -coordinates of process \mathbf{Y} to visit zero (i.e. when $\eta^1(\mathbf{y})$, defined in Step 1, occurs). This estimate is somewhat analogous to a coupon collector's problem (the i -th coupons is collected when the i -th coordinate, Y_i , visits zero).

Assumptions A1) to A3) allow us to obtain suitably uniform estimates on the probability that a particular coupon is collected conditional on the event that a given set of coupons has already been collected. But one has to keep track of the coordinates of the upper bound process each time one attempts to collect a new coupon. We do this by a stochastic domination argument. In the end, we obtain a coupling which implies the bound $\eta^n(\mathbf{y}) \leq \tau^+(\mathbf{y}) + \xi_1 + \dots + \xi_n$ where ξ_i 's are some i.i.d. positive random variables independent of $\tau^+(\mathbf{y})$.

The execution of Step 2.2 requires a number of estimates, but it results in a bound of the following form:

$$E \left[\exp \left(\chi(\theta) \tau^+(\mathbf{y}) + \chi(\theta) \xi_1 \right) \right] \leq \exp \left(h(\mathbf{y}; \theta) \right) E \left[\exp \left(\chi(\theta) \xi_1 \right) \right].$$

Step 2.3 (connecting back to $\mathcal{N}(t; \mathbf{y})$): A standard supermartingale argument, using the domination involving i.i.d. random variables, ξ_i 's, discussed in Step 2.2, results in the bound,

$$E (1 - \beta_0)^{\mathcal{N}(t; \mathbf{y})} = O \left(\exp \left(h(\mathbf{y}; \theta) - \chi(\theta) t \right) \right),$$

which holds uniformly in d as $t \rightarrow \infty$ – assuming that θ is suitably chosen as a function of β_0 . It turns out that the selection of θ forces $\chi(\theta) = O(1/(d^4 \log(d)))$.

Step 3: We conclude the result by putting all of the previous steps together.

3 Step 1: Bounding the Difference of the Coupled Processes

Here, we introduce an auxiliary Markov chain $(W(n) : n \geq 0)$ living on the state space $\{0, 1, \dots, d\}$ so that $P(W(n+1) = j | W(n) = i) = Q_{i,j}$ for $1 \leq i, j \leq d$. State 0 is an absorbing state and $P(W(n+1) = 0 | W(n) = i) = Q_{i,0} = 1 - \sum_{j=1}^d Q_{i,j}$. We use P_i to refer to the probability law given that $W(0) = i$. For any subset $S \subseteq \{1, \dots, d\}$, we define

$$\begin{aligned}\tau(S) &= \inf\{n \geq 0 : W(n) \in S\}, \text{ and} \\ \tau(\{0\}) &= \inf\{n \geq 0 : W(n) = 0\}.\end{aligned}$$

Define the $d \times d$ matrix $\Lambda(S)$ as

$$\Lambda_{i,j}(S) = P_i(\tau(S) < \tau(\{0\}), W(\tau(S)) = j)$$

for $i, j \in \{1, \dots, d\}$.

Lemma 1. *The matrix $\Lambda(S)$ can be represented as*

$$\Lambda(S) = \begin{pmatrix} \Lambda_{SS} & \Lambda_{S\bar{S}} \\ \Lambda_{\bar{S}S} & \Lambda_{\bar{S}\bar{S}} \end{pmatrix} = \begin{pmatrix} I & 0 \\ -(R_{S\bar{S}}R_{\bar{S}\bar{S}}^{-1})^T & 0 \end{pmatrix}.$$

As a result,

$$\Lambda^T(S) = \begin{pmatrix} I & -R_{S\bar{S}}R_{\bar{S}\bar{S}}^{-1} \\ 0 & 0 \end{pmatrix}.$$

Recall that we have defined a sequence of stopping times $\eta_i^k(\mathbf{y})$ and $\eta^k(\mathbf{y})$ in (12). Let

$$\Gamma_i(t, \mathbf{y}) = \{\eta_i^k : \eta_i^k \leq t\}, \text{ and } \Gamma(t, \mathbf{y}) = \cup_{i=1}^d \Gamma_i(t, \mathbf{y}).$$

For any time point $t \geq 0$, define

$$\mathcal{C}(t) = \{1 \leq i \leq d : Y_i(t) = 0\} \text{ and } \bar{\mathcal{C}}(t) = \{1 \leq j \leq d : j \notin \mathcal{C}(t)\}$$

We are ready to provide a bound for $\mathbf{1}^T(\mathbf{Y}(t; \mathbf{y}) - \mathbf{Y}(t; \mathbf{0}))$.

Lemma 2.

$$0 \leq \mathbf{1}^T(\mathbf{Y}(t; \mathbf{y}) - \mathbf{Y}(t; \mathbf{0})) \leq \mathbf{1}^T \prod_{s \in \Gamma(t, \mathbf{y})} \Lambda^T(\bar{\mathcal{C}}(s)) \mathbf{y}.$$

The proofs of Lemma 1 and 2 can be found at the end of this section. Given Lemma 2, we can provide an exponentially decaying upper bound in terms of $\mathcal{N}(t; \mathbf{y})$. The intuition is that the matrices $\Lambda(\bar{\mathcal{C}}(t))$ are substochastic and thus one might hope to obtain an exponentially decaying bound.

Lemma 3.

$$\mathbf{1}^T(\mathbf{Y}(t; \mathbf{y}) - \mathbf{Y}(t; \mathbf{0})) \leq \|\mathbf{y}\|_1 \cdot d\kappa_0 (1 - \beta_0)^{\mathcal{N}(t; \mathbf{y})}.$$

Proof of Lemma 3. For any $k > 0$, we write $\eta_{(1)}^k \leq \eta_{(2)}^k \leq \dots \leq \eta_{(d)}^k$ as the sorting of $\{\eta_1^k, \dots, \eta_d^k\}$. Ties between η_i^k and η_j^k for $i \neq j$ are resolved arbitrarily, for example, lexicographically comparing i and j . For the Markov chain $W(n)$, as we have defined at the beginning of this section, we define a sequence of stopping times τ_j^k as the following:

$$\begin{aligned} \tau_1^1 &= \inf\{n \geq 0 : W(n) \in \bar{\mathcal{C}}(\eta_{(1)}^1)\}, \\ \tau_{j+1}^k &= \inf\{n \geq \tau_j^k : W(n) \in \bar{\mathcal{C}}(\eta_{(j+1)}^k)\} \text{ for all } j \leq d-1, \\ \tau_1^{k+1} &= \inf\{n \geq \tau_d^k : W(n) \in \bar{\mathcal{C}}(\eta_{(1)}^{k+1})\}. \end{aligned}$$

Then, for any $m > 0$ and $1 \leq i \leq d$, one can check that

$$\left(\prod_{k=1}^m \prod_{j=1}^d \Lambda(\bar{\mathcal{C}}(\eta_{(j)}^k)) \mathbf{1} \right)_i = P_i(\tau_1^1 \leq \tau_2^1 \leq \dots \leq \tau_d^m < \tau(\{0\})).$$

We show that $\tau_d^m \geq m$ almost surely conditional on the event that $\tau_1^1 \leq \tau_2^1 \leq \dots \leq \tau_d^m < \tau(\{0\})$. First, we show that $\tau_d^1 \geq 1$. Suppose $\eta_i^1 = \eta_{(j_1)}^1$, then, since $i \notin \bar{\mathcal{C}}(\eta_{(j_1)}^1)$ and $W(0) = i$, we must have $\tau_d^1 \geq \tau_{j_1}^1 \geq 1$. For any $1 \leq k \leq m$, let $l = W(\tau_d^k)$. Suppose $\eta_l^{k+1} = \eta_{(j_l)}^{k+1}$. Since $l \notin \bar{\mathcal{C}}(\eta_{(j_l)}^{k+1})$ and $W(\tau_d^k) = l$, we must have that $\tau_d^{k+1} \geq \tau_{j_l}^{k+1} \geq \tau_d^k + 1$. Therefore, we can conclude by induction

that $\tau_d^m \geq m$, and hence $\tau(\{0\}) \geq m$ conditional on the event that $\tau_1^1 \leq \tau_2^1 \leq \dots \leq \tau_d^m < \tau(\{0\})$.

As a result, we have

$$\left(\prod_{k=1}^m \prod_{j=1}^d \Lambda(\bar{\mathcal{C}}(\eta_{(j)}^k)) \mathbf{1} \right)_i \leq P_i(\tau(\{0\}) \geq m)$$

As $Q_{i,j} = P(W(n+1) = j | W(n) = i)$ for $1 \leq i, j \leq d$ and 0 is the absorbing state,

$$\max_i P_i(\tau(\{0\}) > n) = \|Q^n \mathbf{1}\|_\infty.$$

Under Assumption A1),

$$\|Q^n \mathbf{1}\|_\infty \leq \mathbf{1}^T Q^n \mathbf{1} \leq d \|\mathbf{1}^T Q^n\|_\infty \leq d\kappa_0 (1 - \beta_0)^n.$$

As a result, we have

$$\left\| \prod_{k=1}^m \prod_{j=1}^d \Lambda(\bar{\mathcal{C}}(\eta_{(j)}^k)) \mathbf{1} \right\|_\infty \leq d\kappa_0 (1 - \beta_0)^m.$$

Let $t_1 = \eta^{\mathcal{N}(t; \mathbf{y})}$ and recall from the definition of $\mathcal{N}(t; \mathbf{y})$ that $t_1 \leq t$. Then, we have

$$\begin{aligned} \mathbf{1}^T (\mathbf{Y}(t; \mathbf{y}) - \mathbf{Y}(t; \mathbf{0})) &\leq \mathbf{1}^T (\mathbf{Y}(t_1; \mathbf{y}) - \mathbf{Y}(t_1; \mathbf{0})) \\ &\leq \mathbf{1}^T \prod_{k=1}^{\mathcal{N}(t; \mathbf{y})} \prod_{j=1}^d \Lambda^T(\bar{\mathcal{C}}(\eta_{(j)}^k)) \mathbf{y} \\ &\leq \left\| \prod_{k=1}^{\mathcal{N}(t; \mathbf{y})} \prod_{j=1}^d \Lambda(\bar{\mathcal{C}}(\eta_{(j)}^k)) \mathbf{1} \right\|_\infty \|\mathbf{y}\|_1 \\ &\leq \|\mathbf{y}\|_1 d\kappa_0 (1 - \beta_0)^{\mathcal{N}(t; \mathbf{y})}. \end{aligned}$$

Here, the first inequality follows Theorem 1 of [7] and the second inequality follows Lemma 2. \square

Proof of Lemma 1. Following the definition of the matrix $\Lambda(S)$, it is obvious that, for all $j \in \bar{S}$,

$$\Lambda_{i,j}(S) = 0 \text{ as } P(W(\tau(S)) = j) = 0,$$

and for all $i, j \in S$,

$$\Lambda_{i,j}(S) = \delta_{i,j} \text{ as } \tau(S) = 0 \text{ and } W(\tau(S)) = i.$$

Therefore, $\Lambda_{SS} = I$ and all elements of $\Lambda_{S\bar{S}}$ and $\Lambda_{\bar{S}\bar{S}}$ are 0. By the property of Markov chains with transient states, we can compute that

$$\begin{aligned}\Lambda_{\bar{S}S} &= Q_{\bar{S}S} + Q_{\bar{S}\bar{S}}Q_{\bar{S}S} + Q_{\bar{S}\bar{S}}^2Q_{\bar{S}S} + \dots \\ &= (I + Q_{\bar{S}\bar{S}} + Q_{\bar{S}\bar{S}}^2 + \dots)Q_{\bar{S}S} = (I - Q_{\bar{S}\bar{S}})^{-1}Q_{\bar{S}S}\end{aligned}$$

Note that $R = (I - Q)^T$. As a result, we have that $(I - Q_{\bar{S}\bar{S}}) = R_{\bar{S}\bar{S}}^T$ and $Q_{\bar{S}S} = -R_{\bar{S}\bar{S}}^T$, and therefore $\Lambda_{\bar{S}S} = -(R_{\bar{S}\bar{S}}R_{\bar{S}\bar{S}}^{-1})^T$. \square

Proof of Lemma 2. For simplicity of notation, we write $\tilde{\mathbf{Y}}(t) = \mathbf{Y}(t; \mathbf{y})$ and $\mathbf{Y}(t) = \mathbf{Y}(t; \mathbf{0})$. Since $\Gamma(t, \mathbf{y})$ is a finite set for all t , let t_1 be the maximum of set $\Gamma(t, \mathbf{y})$ and denote $\mathcal{C} = \mathcal{C}(t_1)$. If $\Gamma(t, \mathbf{y})$ is empty, we define $t_1 = 0$. We will prove the following statement:

$$\tilde{\mathbf{Y}}(t_1) - \mathbf{Y}(t_1) \leq \prod_{s \in \Gamma(t, \mathbf{y})} \Lambda^T(\bar{\mathcal{C}}(s)) \mathbf{y} + H\mathbf{w}, \quad (14)$$

for some $\mathbf{w} \geq 0$ and H is a matrix defined via

$$H_{ij} = 1(i \in \bar{\mathcal{C}}, j \in \bar{\mathcal{C}}) \cdot (P_j(\tau(\{i\}) < \tau(\{0\})) \text{ and } W(n) \in \mathcal{C} \text{ for all } n \leq \tau(\{i\}) - 1) - \delta_{ij}.$$

Then, we can conclude

$$\mathbf{1}^T(\tilde{\mathbf{Y}}(t) - \mathbf{Y}(t)) \leq \mathbf{1}^T(\tilde{\mathbf{Y}}(t_1) - \mathbf{Y}(t_1)) \leq \mathbf{1}^T \prod_{s \in \Gamma(t, \mathbf{y})} \Lambda^T(\bar{\mathcal{C}}(s)) \mathbf{y},$$

where the first inequality holds following Part (iv) of Theorem 1 in [7] and the last holds as $\mathbf{1}^T H \leq 0$.

Now, we shall prove (14) by induction on the cardinality of $\Gamma(t, \mathbf{y})$. The base case is that $\Gamma(t, \mathbf{y})$ is empty. Then, for any t , as long as $\Gamma(t, \mathbf{y})$ is empty, $t_1 = 0$ and hence

$$\tilde{\mathbf{Y}}(t_1) - \mathbf{Y}(t_1) = \tilde{\mathbf{Y}}(0) - \mathbf{Y}(0) = \mathbf{y}$$

and (14) holds for $\mathbf{w} = 0$.

Suppose (14) holds for all t such that the cardinality of $\Gamma(t, \mathbf{y}) \leq k$. Consider the case that $\Gamma(t, \mathbf{y}) = k + 1$. Let t_2 be the second largest element of the set $\Gamma(t, \mathbf{y})$. Let $\mathbf{z} = \tilde{\mathbf{Y}}(t_2) - \mathbf{Y}(t_2)$ and

$\mathbf{w} = (\mathbf{L}(t_1) - \mathbf{L}(t_2)) - (\tilde{\mathbf{L}}(t_1) - \tilde{\mathbf{L}}(t_2)) \geq 0$ (see Theorem 1 in [7]). At time t_1 , by definition, we have

$$\tilde{\mathbf{Y}}(t_1) - \mathbf{Y}(t_1) = \mathbf{z} - R\mathbf{w}.$$

As $\tilde{\mathbf{Y}}_{\mathcal{C}}(t_1) = \mathbf{Y}_{\mathcal{C}}(t_1) = 0$,

$$0 = \tilde{\mathbf{Y}}_{\mathcal{C}}(t_1) - \mathbf{Y}_{\mathcal{C}}(t_1) = \mathbf{z}_{\mathcal{C}} - R_{\mathcal{C}\mathcal{C}}\mathbf{w}_{\mathcal{C}} - R_{\mathcal{C}\bar{\mathcal{C}}}\mathbf{w}_{\bar{\mathcal{C}}},$$

from which we solve $\mathbf{w}_{\mathcal{C}} = R_{\mathcal{C}\mathcal{C}}^{-1}(\mathbf{z}_{\mathcal{C}} - R_{\mathcal{C}\bar{\mathcal{C}}}\mathbf{w}_{\bar{\mathcal{C}}})$. Therefore,

$$\begin{aligned} \tilde{\mathbf{Y}}_{\bar{\mathcal{C}}}(t_1) - \mathbf{Y}_{\bar{\mathcal{C}}}(t_1) &= \mathbf{z}_{\bar{\mathcal{C}}} - R_{\bar{\mathcal{C}}\mathcal{C}}\mathbf{w}_{\mathcal{C}} - R_{\bar{\mathcal{C}}\bar{\mathcal{C}}}\mathbf{w}_{\bar{\mathcal{C}}} \\ &= \mathbf{z}_{\bar{\mathcal{C}}} - R_{\bar{\mathcal{C}}\mathcal{C}}R_{\mathcal{C}\mathcal{C}}^{-1}(\mathbf{z}_{\mathcal{C}} - R_{\mathcal{C}\bar{\mathcal{C}}}\mathbf{w}_{\bar{\mathcal{C}}}) - R_{\bar{\mathcal{C}}\bar{\mathcal{C}}}\mathbf{w}_{\bar{\mathcal{C}}} \\ &= (I\mathbf{z}_{\bar{\mathcal{C}}} - R_{\bar{\mathcal{C}}\mathcal{C}}R_{\mathcal{C}\mathcal{C}}^{-1}\mathbf{z}_{\mathcal{C}}) + (R_{\bar{\mathcal{C}}\mathcal{C}}R_{\mathcal{C}\mathcal{C}}^{-1}R_{\mathcal{C}\bar{\mathcal{C}}} - R_{\bar{\mathcal{C}}\bar{\mathcal{C}}})\mathbf{w}_{\bar{\mathcal{C}}} \\ &= \Lambda_{\bar{\mathcal{C}}}^T(\bar{\mathcal{C}})\mathbf{z} + (R_{\bar{\mathcal{C}}\mathcal{C}}R_{\mathcal{C}\mathcal{C}}^{-1}R_{\mathcal{C}\bar{\mathcal{C}}} - R_{\bar{\mathcal{C}}\bar{\mathcal{C}}})\mathbf{w}_{\bar{\mathcal{C}}}, \end{aligned}$$

where the last equation holds following Lemma 1. Note that

$$R_{\bar{\mathcal{C}}\mathcal{C}}R_{\mathcal{C}\mathcal{C}}^{-1}R_{\mathcal{C}\bar{\mathcal{C}}} - R_{\bar{\mathcal{C}}\bar{\mathcal{C}}} = Q_{\bar{\mathcal{C}}\mathcal{C}}^T(I - Q_{\mathcal{C}\mathcal{C}}^T)^{-1}Q_{\mathcal{C}\bar{\mathcal{C}}}^T + Q_{\bar{\mathcal{C}}\bar{\mathcal{C}}}^T - I_{\bar{\mathcal{C}}},$$

where Q is the transition matrix of W . Let $H_{\bar{\mathcal{C}}\bar{\mathcal{C}}} = R_{\bar{\mathcal{C}}\mathcal{C}}R_{\mathcal{C}\mathcal{C}}^{-1}R_{\mathcal{C}\bar{\mathcal{C}}} - R_{\bar{\mathcal{C}}\bar{\mathcal{C}}}$. From the definition of Q , we can check that

$$H_{ij} = (P_j(\tau_i < \tau(\{0\}) \text{ and } W(n) \in \mathcal{C} \text{ for all } n \leq \tau_i - 1) - \delta_{ij}),$$

for all $i, j \in \bar{\mathcal{C}}$ and $\tau_i := \inf\{n \geq 1 : W(n) = i\}$. Note that $\Lambda_{\mathcal{C}}(\bar{\mathcal{C}}) = 0$ following Lemma 1, so we have

$$\tilde{\mathbf{Y}}(t_1) - \mathbf{Y}(t_1) = \Lambda^T(\mathcal{C})\mathbf{z} + H\mathbf{w}.$$

Note that the cardinality of $\Gamma(t_2, \mathbf{y}) = k$ and t_2 is its maximum, so by induction, we have

$$\mathbf{z} \leq \prod_{s \in \Gamma(t, \mathbf{y}) \setminus \{t_1\}} \Lambda^T(\bar{\mathcal{C}}(s))\mathbf{y} + H^*\mathbf{w}^*,$$

where $\mathbf{w}^* \geq 0$ and

$$H_{ij}^* = 1(i \in \bar{D}, j \in \bar{D})(P_j(\tau_i < \tau(\{0\}) \text{ and } W(n) \in D \text{ for all } n \leq \tau_i - 1) - \delta_{ij}),$$

with $D = \mathcal{C}(t_2)$. As $\Lambda(\bar{\mathcal{C}}) \geq 0$, so we have

$$\tilde{\mathbf{Y}}(t_1) - \mathbf{Y}(t_1) \leq \prod_{s \in \Gamma(t, \mathbf{y})} \Lambda^T(\bar{\mathcal{C}}(s)) \mathbf{y} + \Lambda^T(\bar{\mathcal{C}}) H^* \mathbf{w}^* + H \mathbf{w}.$$

As $\mathbf{w}^* \geq 0$, it suffices to show that $(\Lambda^T(\bar{\mathcal{C}}) H^*)_{ij} \leq 0$ for all $1 \leq i, j \leq d$. Note that

$$(\Lambda^T(\bar{\mathcal{C}}) H^*)_{ij} = \sum_k \Lambda^T(\bar{\mathcal{C}})_{ik} H_{kj}^*.$$

Since $H_{kj}^* = 0$ for all $j \in D$, we conclude that $(\Lambda^T(\bar{\mathcal{C}}) H^*)_{ij} = 0$ for all $j \in D$.

For $j \in \bar{D}$, recall that $\Lambda^T(\bar{\mathcal{C}})_{ij} = P_j(\tau(\bar{\mathcal{C}}) < \tau(\{0\}), W(\tau(\bar{\mathcal{C}})) = i)$, therefore

$$\begin{aligned} (\Lambda^T(\bar{\mathcal{C}}) H^*)_{ij} &= \sum_k \Lambda^T(\bar{\mathcal{C}})_{ik} H_{kj}^* \\ &= \sum_{k \in \bar{D}} P_k(\tau(\bar{\mathcal{C}}) < \tau(\{0\}), W(\tau(\bar{\mathcal{C}})) = i) (P_j(\tau_k < \tau(\{0\}) \text{ and } W(n) \in D \text{ for all } n \leq \tau_k - 1) - \delta_{kj}) \\ &= P_j(\tau(\bar{D}) < \tau(\{0\}), \tilde{\tau}(\bar{\mathcal{C}}) < \tau(\{0\}), W(n) \in D \text{ for all } n < \tau(\bar{D}), W(\tau(\bar{\mathcal{C}})) = i) \\ &\quad - P_j(\tau(\bar{\mathcal{C}}) < \tau(\{0\}), W(\tau(\bar{\mathcal{C}})) = i) \\ &\leq 0, \end{aligned} \tag{15}$$

where $\tilde{\tau}(\bar{\mathcal{C}}) \doteq \inf\{t \geq \tau(\bar{D}) : W(t) \in \bar{\mathcal{C}}\}$ and the inequality holds as the first probability event is a subset of the latter one in (15). \square

4 Step 2: Coupling, Lyapunov Bounds, and Geometric Trials

One of the main results in this section is the following.

Proposition 1. *Under A1) to A3), for any $\beta > 0$ satisfying (9), we have*

$$E \left[(1 - \beta)^{\mathcal{N}(t; \mathbf{y})} \right] \leq \exp(\zeta_0 \|\mathbf{y}\|_\infty / (d^3 \log(d)) + \beta/d^2) \cdot \exp(-\zeta_1 t / (d^4 \log(d))) \cdot (1 - \beta)^{-1},$$

where

$$\zeta_0 = \frac{\delta_1 \cdot \beta^2}{2 \max_{i=1}^d \sigma_i^2}, \quad \zeta_1 = \frac{\delta_1^2 \cdot \beta^2}{16 \max_{i=1}^d \sigma_i^2}.$$

The proof of Proposition 1 follows Steps 2.1, 2.2 and 2.3 as described in the main strategy. The proofs of all the technical lemmas can be found in Section 4.1.

We first explain how to construct the upper bound process $\mathbf{Y}^+(\cdot; \mathbf{y})$ briefly mentioned in the discussion of Step 2. Following Assumptions A1) and A3), $\|R^{-1}\mathbf{1}\|_\infty \leq \kappa_0/\beta_0$, and $R^{-1}\boldsymbol{\mu} \leq -\delta_0\mathbf{1}$. We choose

$$\boldsymbol{\mu}^+ = \boldsymbol{\mu} + \delta_1\mathbf{1},$$

where $\delta_1 = \delta_0\beta_0/(2\kappa_0)$. One can check that $\boldsymbol{\mu}^+ > \boldsymbol{\mu}$ and $R^{-1}\boldsymbol{\mu}^+ \leq -(\delta_0/2)\mathbf{1}$.

Let $(\mathbf{Y}^+(\cdot), \mathbf{L}^+(\cdot))$ be the solution to the Skorokhod problem with orthogonal reflection as follows,

$$\mathbf{Y}^+(t) = \mathbf{Y}^+(0) + \bar{\mathbf{X}}(t) + \mathbf{L}^+(t),$$

with $\bar{\mathbf{X}}(t) = \mathbf{X}(t) - \boldsymbol{\mu}^+t$ and $\mathbf{Y}^+(0) = \mathbf{y}$. We write $\mathbf{Y}^+(t)$ as $\mathbf{Y}^+(t; \mathbf{y})$, as its value depends on the initial value \mathbf{y} . We know from Lemma 3.1 in [8] that

$$R^{-1}\mathbf{Y}(t; \mathbf{y}) \leq R^{-1}\mathbf{Y}^+(t; \mathbf{y}). \quad (16)$$

As discussed in Step 2.1, we have defined $\tau^+(\mathbf{y}) = \inf\{t \geq 0 : \mathbf{Y}^+(t; \mathbf{y}) \leq \mathbf{1}\}$, which is the time to visit a compact set for \mathbf{Y}^+ , and for \mathbf{Y} as well, according to (16):

$$\mathbf{Y}(\tau^+(\mathbf{y}); \mathbf{y}) \leq R^{-1}\mathbf{Y}(\tau^+(\mathbf{y}); \mathbf{y}) \leq R^{-1}\mathbf{Y}^+(\tau^+(\mathbf{y}); \mathbf{y}) \leq R^{-1}\mathbf{1} \leq \frac{\kappa_0}{\beta_0}\mathbf{1} = b_1\mathbf{1},$$

where the first inequality holds as $R^{-1} \geq I$ and $\mathbf{Y} \geq 0$. The following result provides a bound for the moment-generating function of $\tau^+(\mathbf{y})$.

Lemma 4. *Define*

$$g(y) = 2^{-1}y^2I(0 \leq y \leq 1) + (y - 1/2)I(y > 1).$$

For any given $\varepsilon > 0$ and $\theta > 0$, define

$$h(\mathbf{y}; \theta) = \varepsilon \log \left(\sum_{i=1}^d \exp(g(\theta y_i)/\varepsilon) \right) \leq \max_{i=1}^d g(\theta y_i) + \varepsilon \log(d) \leq \theta \|\mathbf{y}\|_\infty + \varepsilon \log(d).$$

Then, for any

$$0 < \theta \leq \frac{\varepsilon}{2\varepsilon + 1} \cdot \frac{\delta_1}{(1+d) \max_{i=1}^d \sigma_i^2} \leq \frac{\varepsilon}{2\varepsilon + 1} \cdot \frac{\delta_0 \beta_0 / (2\kappa_0)}{(1+d) b_0},$$

and

$$\chi(\theta) \doteq \theta \frac{\delta_1}{2(1+d)} \leq \theta \frac{\delta_0 \beta_0 / (2\kappa_0)}{2(1+d)}, \quad (17)$$

we have

$$E \left[\exp(h(\mathbf{Y}^+(\tau^+(\mathbf{y})); \theta) + \chi(\theta) \tau^+(\mathbf{y})) \right] \leq \exp(h(\mathbf{y}; \theta)) \leq \exp(\theta \|\mathbf{y}\|_\infty + \varepsilon \log(d)). \quad (18)$$

Starting from position $\mathbf{Y}(\tau^+(\mathbf{y}))$, we wait for another unit of time till $\tau^+(\mathbf{y}) + 1$. If the event $\{Y_i(t) = 0 \text{ for some } \tau^+(\mathbf{y}) < t \leq \tau^+(\mathbf{y}) + 1\}$ occurs, then we can conclude that $\eta_i^1 \leq \tau^+(\mathbf{y}) + 1$. The following lemma shows that, for all $1 \leq i \leq d$, the probability for such an event to happen is uniformly bounded away from 0, regardless of the position of the process at time $\tau^+(\mathbf{y})$.

Lemma 5. *There exists a constant $p_0 > 0$, independent of d , such that*

$$p_0 \leq P \left(N(0, 1) < \sqrt{b_0}(\delta_0 - b_1^2) \right). \quad (19)$$

Besides, for all $\mathbf{y} \leq b_1 \mathbf{1}$ and every $i \in \{1, \dots, d\}$

$$P(Y_i(t) = 0 \text{ for some } t \leq 1 | \mathbf{Y}(0) = \mathbf{y}) \geq p_0.$$

Based on Lemma 5, we are ready to perform a “geometric trial argument” (Step 2.2) to obtain a bound for each η_i^1 with $1 \leq i \leq d$. Each round of the trials includes two steps described as follows. Suppose at the beginning of the k -th round of trial, the initial position of the process \mathbf{Y} is $\mathbf{Y}^{i,k}$ (in particular, $\mathbf{Y}^{i,1} = \mathbf{y}$). In the first step, it takes $\tau^+(k; \mathbf{Y}^{i,k})$ for $\mathbf{Y}(\cdot; \mathbf{Y}^{i,k})$ to arrive to the compact set $\{\mathbf{Y} \in \mathbb{R}^d : |y_i| \leq b_1\}$. (For given \mathbf{y} , $\tau^+(k; \mathbf{y})$'s are i.i.d. copies of $\tau^+(\mathbf{y})$.) Then, in the next one unit of time, we check if the event $\{Y_i(t; \mathbf{Y}^{i,k}) = 0 \text{ for some } \tau^+(k; \mathbf{Y}^{i,k}) < t \leq \tau^+(k; \mathbf{Y}^{i,k}) + 1\}$

happens. If so, we can stop as the process has already hit 0. If not, we then start the next round of trial with the initial position $\mathbf{Y}^{i,k+1} = \mathbf{Y}(\tau^+(k; \mathbf{Y}^{i,k}) + 1; \mathbf{Y}^{i,k})$. In summary, we can define a sequence of Bernoulli random variables $\zeta_k(i)$ jointly with the sequence $\{\mathbf{Y}^{i,k}\}$ as

$$\zeta_k(i) \doteq 1(Y_i(t; \mathbf{Y}^{i,k}) = 0 \text{ for some } \tau^+(k; \mathbf{Y}^{i,k}) < t \leq \tau^+(k; \mathbf{Y}^{i,k}) + 1).$$

Let $K = \min\{k : \zeta_k(i) = 1\}$, and we obtain a bound for $\eta_i^1(\mathbf{y})$:

$$\eta_i^1(\mathbf{y}) \leq \sum_{k=1}^K (\tau^+(k; \mathbf{Y}^{i,k}) + 1).$$

The next lemma shows that we can replace K with a Geometric random variable (r.v.) G^i , and the sequence $\tau^+(\mathbf{Y}^{i,k})$ with an i.i.d. sequence of positive r.v.'s that are independent of G and have bounded moment-generating function.

Lemma 6. *Let p be any positive number such that $p < p_0$. Let $\{G^i : 1 \leq i \leq d\}$ be i.i.d. copies of a Geometric random variable G with probability of success equal to p . Then, we can construct a random variable $\Theta_d > 0$ and its i.i.d. copies $\{\Theta_d^{i,k}\}$ such that*

$$\eta_i^1(\mathbf{y}) \leq \tau^+(\mathbf{y}) + \sum_{k=1}^{G^i} \left(1 + \tau^+(\Theta_d^{i,k} \mathbf{1})\right).$$

Therefore,

$$\eta^1(\mathbf{y}) \leq \tau^+(\mathbf{y}) + \sum_{i=1}^d \sum_{k=1}^{G^i} \left(1 + \tau^+(\Theta_d^{i,k} \mathbf{1})\right).$$

Moreover, let $\phi_d(\theta) = E[\exp(\theta \Theta_d)]$, then, for $\theta = o(1)$ as $d \rightarrow \infty$,

$$\phi_d(\theta) \leq 1 + 2(1-p)^{-1} \theta \log(1+d) \exp\left(\theta \log(1+d)^{2/3}\right) + \theta O\left(\exp\left(-\log(1+d)^{4/3}/3b_0\right)\right). \quad (20)$$

Define a random variable

$$\xi = \sum_{i=1}^d \sum_{k=1}^{G^i} \left(1 + \tau^+(\Theta_d^{i,k} \mathbf{1})\right).$$

According to Lemma 6, we can couple $\eta^1(\mathbf{y})$ and ξ so that

$$\eta^1(\mathbf{y}) \leq \tau^+(\mathbf{y}) + \xi,$$

where $\tau^+(\mathbf{y})$ is independent of ξ . The Skorokhod problem is monotone with respect to the initial condition, i.e. $\eta^1(\mathbf{y}) \leq \eta^1(\mathbf{y}')$ whenever $\mathbf{y} \leq \mathbf{y}'$. As a result, we can iteratively apply the previous reasoning. In particular, let ξ_1, ξ_2, \dots be iid copies of ξ and independent of $\tau^+(\mathbf{y})$. Then, we can construct a coupling so that

$$\begin{aligned} \eta^1(\mathbf{y}) &\leq \tau^+(\mathbf{y}) + \xi_1, \\ \eta^2(\mathbf{y}) &\leq \tau^+(\mathbf{y}) + \xi_1 + \xi_2, \\ &\dots \\ \eta^n(\mathbf{y}) &\leq \tau^+(\mathbf{y}) + \xi_1 + \dots + \xi_n. \end{aligned} \tag{21}$$

Based on the bound of the moment-generating function of $\tau^+(\mathbf{y})$ in Lemma 4 and Lemma 6, we have the following result on the moment-generating function of $\eta^n(\mathbf{y})$ for all $n \geq 1$.

Lemma 7. *For $n \geq 1$,*

$$E \exp(\chi(\theta) \eta^n(\mathbf{y})) \leq \exp(h(\mathbf{y}; \theta)) \left(\frac{\phi_d(\theta) \exp(\chi(\theta) + \varepsilon \log(d)) p}{1 - (1-p)\phi_d(\theta) \exp(\chi(\theta) + \varepsilon \log(d))} \right)^{nd}.$$

Moreover, suppose that $\varepsilon, \theta > 0$ are chosen so that

$$\phi_d(\theta) \exp(\chi(\theta) + \varepsilon \log(d)) \leq \frac{1}{(1-p)(1+p)}. \tag{22}$$

Then,

$$E \exp(\chi(\theta) \eta^n(\mathbf{y})) \leq \exp(h(\mathbf{y}; \theta)) (1-p)^{-nd}.$$

Finally, we obtain the following lemma, which takes us very close to the proof of Proposition 1.

Lemma 8. *Assume that $p = \min(p_0, \beta/d)$, ε and $\theta > 0$ satisfies (22) and $\chi(\theta)$ is defined according*

to (17), we obtain that

$$E \left((1-p)^{d \cdot \mathcal{N}(t; \mathbf{y})} \right) \leq \exp(h(\mathbf{y}; \theta)) \cdot \exp(-\chi(\theta)t) \cdot (1-p)^{-d}.$$

We now have all the ingredients required to provide a the proof of Proposition 1.

Proof of Proposition 1. By Lemma 8, the only step that remains is to select θ, ϵ satisfying (22) and to estimate the behavior of $\chi(\theta)$ assuming our selection of p in Lemma 8. Given that $p = \min(p_0, \beta/d)$, we have

$$(1-p)^d \geq (1-\beta).$$

We then choose ϵ, θ as follows:

$$\begin{aligned} \epsilon &= \frac{\beta^2}{2d^2 \log(d)}, \\ \theta &= \frac{\epsilon}{2\epsilon + 1} \cdot \frac{\delta_1}{(1+d) \max_{i=1}^d \sigma_i^2} \leq \frac{\delta_1 \cdot \beta^2}{2d^3 \log(d) \max_{i=1}^d \sigma_i^2}, \end{aligned}$$

and hence

$$\chi(\theta) = \theta \frac{\delta_1}{2(1+d)} \leq \frac{\delta_1^2 \cdot \beta^2}{4d^4 \log(d) \max_{i=1}^d \sigma_i^2}.$$

Therefore, for d sufficiently large,

$$\begin{aligned} &\phi_d(\theta) \exp(\chi(\theta) + \epsilon \log(d)) \\ &\leq \exp(\chi(\theta) + \epsilon \log(d)) \left(1 + 2(1-p)^{-1} \theta \log(1+d) \exp\left(\theta \log(2+d)^{2/3}\right) + \frac{\epsilon}{4} \right) \\ &\leq \exp\left(\frac{\delta_1^2 \cdot \beta^2}{4d^4 \log(d) \max_{i=1}^d \sigma_i^2} + \frac{\beta^2}{2d^2}\right) \left(1 + 2(1-p)^{-1} \theta \log(1+d) \exp\left(\theta \log(1+d)^{2/3}\right) + \frac{\epsilon}{4} \right), \end{aligned}$$

where the first inequality follows from (20) and the fact that the big-O term in (20) goes to 0 as $d \rightarrow \infty$. Given our choice of θ , we have

$$\exp\left(\frac{\delta_1^2 \cdot \beta^2}{4d^4 \log(d) \max_{i=1}^d \sigma_i^2}\right) = 1 + o(d^{-2}), \text{ and } \theta \log(1+d) \exp\left(\theta \log(1+d)^{2/3}\right) = o(d^{-2}),$$

as $d \rightarrow \infty$. Hence, our choice of ε and θ satisfies that, for d sufficiently large,

$$\begin{aligned}\phi_d(\theta) \exp(\chi(\theta) + \varepsilon \log(d)) &\leq \left(1 + \frac{\beta^2}{2d^2}\right) \left(1 + \frac{\beta^2}{4d^2}\right) + o(d^{-2}) \\ &\leq 1 + \frac{\beta^2}{d^2} \leq \frac{1}{(1 - \beta/d)(1 + \beta/d)},\end{aligned}$$

which is exactly the inequality (22). On the other hand, note that $\epsilon \leq 1/2$, so when $d \geq 3$, we have

$$\chi(\theta) = \theta \frac{\delta_1}{2(1+d)} = \frac{1}{2\varepsilon + 1} \cdot \frac{\delta_1^2 \beta^2}{4d^2(1+d)^2 \log(d) \max_{i=1}^d \sigma_i^2} \geq \frac{\delta_1^2 \beta^2}{16d^4 \log(d) \max_{i=1}^d \sigma_i^2}.$$

Now, let

$$\zeta_0 = \frac{\delta_1 \cdot \beta^2}{2 \max_{i=1}^d \sigma_i^2}, \quad \zeta_1 = \frac{\delta_1^2 \cdot \beta^2}{16 \max_{i=1}^d \sigma_i^2}.$$

According to Lemma 8 and the fact that $(1-p)^d \geq (1-\beta/d)^d \geq 1-\beta$, we have

$$\begin{aligned}E(1-\beta)^{\mathcal{N}(t;\mathbf{y})} &\leq \exp(h(\mathbf{y}; \theta) \exp(-\chi(\theta)t)(1-p)^{-d}) \\ &\leq \exp(\theta \|\mathbf{y}\|_\infty + \varepsilon \log(d)) \exp(-\chi(\theta)t)(1-p)^{-d} \\ &\leq \exp(\zeta_0 \|\mathbf{y}\|_\infty / (d^3 \log(d)) + \beta/d^2) \cdot \exp(-\zeta_1 t / (d^4 \log(d))) \cdot (1-\beta)^{-1},\end{aligned}$$

where the second inequality follows Lemma 4 and the last inequality follows our choice of θ and ε . □

We close this section with the proof of the technical results behind the proof of Proposition 1.

4.1 Technical Proofs of Auxiliary Results Behind Proposition 1

We provide the proofs in the order in which we presented the auxiliary results. First, the main ingredient behind Lemma 4 is the following result:

Lemma 9. *Suppose that there exists a non-negative function $h(\cdot)$ and a constant $\chi > 0$ satisfying the following two conditions:*

1. For all $\mathbf{y} = (y_1, \dots, y_d)^T \in R_+^d$ with $\|\mathbf{y}\|_\infty \geq b$

$$(\boldsymbol{\mu} - \boldsymbol{\mu}^+)^T Dh(\mathbf{y}) + \frac{1}{2}Tr(\Sigma D^2h(\mathbf{y})) + \frac{1}{2}Dh(\mathbf{y})^T \Sigma Dh(\mathbf{y}) \leq -\chi, \quad (23)$$

where $Dh(\mathbf{y})$ and $D^2h(\mathbf{y})$ are the first and second derivatives of $h(\cdot)$ evaluated at \mathbf{y} , respectively. (We encode $Dh(\mathbf{y})$ as column vector.)

2. For any $\mathbf{y} = (y_1, \dots, y_d)^T \in \partial R_+^d$,

$$Dh(\mathbf{y})^T \mathbf{w} \leq 0 \text{ for all } \mathbf{w} \in \mathcal{Z}_{\mathbf{y}}, \quad (24)$$

where

$$\mathcal{Z}_{\mathbf{y}} = \{\mathbf{w} = (w_1, \dots, w_d)^T \in R_+^d : w_l > 0 \text{ if and only if } y_l = 0\}.$$

Then, for any $\|\mathbf{y}\|_\infty \geq 1$,

$$E \exp(h(\mathbf{Y}^+(\tau^+(\mathbf{y}))) + \chi \tau^+(\mathbf{y})) \leq \exp(h(\mathbf{y})).$$

In particular,

$$E \exp(\chi \tau^+(\mathbf{y})) \leq \exp(h(\mathbf{y})).$$

Proof of Lemma 9. Note that Ito's lemma yields that for a twice continuously differentiable $h(\cdot)$

$$\begin{aligned} & h(\mathbf{Y}^+(t)) - h(\mathbf{Y}^+(0)) \\ &= \int_0^t (\mathcal{A}h)(\mathbf{Y}^+(s)) ds + \int_0^t Dh(\mathbf{Y}^+(s)) d\mathbf{L}^+(s) + \int_0^t Dh(\mathbf{Y}^+(s)) C d\mathbf{B}(s), \end{aligned} \quad (25)$$

where C is the Cholesky decomposition matrix such that $CC^T = \Sigma$, and

$$(\mathcal{A}h)(\mathbf{y}) ds \doteq (\boldsymbol{\mu} - \boldsymbol{\mu}^+)^T Dh(\mathbf{y}) + \frac{1}{2}Tr(\Sigma D^2h(\mathbf{y})).$$

We know that

$$\bar{M}(t) = \exp \left(\int_0^t Dh(\mathbf{Y}(s)) C d\mathbf{B}(s) - \frac{1}{2} \int_0^t Dh(\mathbf{Y}(s))^T \Sigma Dh(\mathbf{Y}(s)) ds \right)$$

is a non-negative local martingale and, therefore, a supermartingale. We thus conclude that

$$E_{\mathbf{y}} \bar{M}(t) \leq 1.$$

Substituting (25) into $\bar{M}(t)$ and using the assumptions on $h(\cdot)$, we obtain that

$$E_{\mathbf{y}} \exp(h(\mathbf{Y}^+(\tau^+(\mathbf{y}))) - h(\mathbf{y}) + \chi \tau^+(\mathbf{y})) \leq E_{\mathbf{y}} \bar{M}(t) \leq 1.$$

Because $h(\cdot) \geq 0$, we conclude that

$$E_{\mathbf{y}} \exp(-h(\mathbf{y}) + \chi \tau^+(\mathbf{y})) \leq 1,$$

which is equivalent to the statement of the result. □

Using the previous result, we now can provide the proof of Lemma 4.

Proof of Lemma 4. We start by computing the first and second derivatives of $h(\cdot)$. Let

$$w_i(y, \varepsilon) = \frac{\exp(g(\theta y_i)/\varepsilon)}{\sum_{j=1}^d \exp(g(\theta y_j)/\varepsilon)}.$$

Note that

$$\begin{aligned} Dh(\mathbf{y}) &= \sum_{i=1}^d w_i(y, \varepsilon) g'(\theta e_i^T \mathbf{y}) \theta e_i \\ D^2 h(\mathbf{y}) &= \theta^2 \sum_{i=1}^d w_i(y, \varepsilon) g''(\theta e_i^T \mathbf{y}) e_i e_i^T \\ &\quad + \frac{\theta^2}{\varepsilon} \sum_{i=1}^d w_i(y, \varepsilon) g'(\theta e_i^T \mathbf{y})^2 e_i e_i^T \\ &\quad - \frac{\theta^2}{\varepsilon} \sum_{i,j=1}^d w_i(y, \varepsilon) w_j(y, \varepsilon) g'(\theta e_i^T \mathbf{y}) g'(\theta e_j^T \mathbf{y}) e_i e_j^T. \end{aligned}$$

Therefore,

$$\begin{aligned}
Tr(\Sigma D^2 h(\mathbf{y})) &\leq \frac{\theta^2}{\varepsilon} \sum_{i=1}^d w_i(y, \varepsilon) \sigma_i^2 (\varepsilon g''(\theta e_i^T \mathbf{y}) + g'(\theta e_i^T \mathbf{y})^2), \\
Dh(\mathbf{y})^T \Sigma Dh(\mathbf{y}) &= \theta^2 \left(\sum_{i=1}^d w_i(y, \varepsilon) g'(\theta e_i^T \mathbf{y}) \sigma_i \right)^2 \leq \theta^2 \max_{i=1}^d \sigma_i^2, \\
(\boldsymbol{\mu} - \boldsymbol{\mu}^+)^T Dh(\mathbf{y}) &= -\theta \sum_{i=1}^d w_i(y, \varepsilon) g'(\theta e_i^T \mathbf{y}) \delta_1.
\end{aligned}$$

Because $-w_i(y, \varepsilon) g'(\theta e_i^T \mathbf{y}) \delta_1 \leq 0$, we have that

$$\begin{aligned}
-\sum_{i=1}^d w_i(y, \varepsilon) g'(\theta e_i^T \mathbf{y}) \delta_1 &\leq -\sum_{i=1}^d w_i(y, \varepsilon) g'(\theta e_i^T \mathbf{y}) \delta_1 I(e_i^T \mathbf{y} \geq 1) \\
&= -\sum_{i=1}^d w_i(y, \varepsilon) \delta_1 I(e_i^T \mathbf{y} \geq 1) \\
&\leq \frac{-\delta_1}{1+d},
\end{aligned}$$

where in the last inequality we use the fact that, for $\|\mathbf{y}\|_\infty \geq 1$,

$$\sum_{i=1}^d w_i(y, \varepsilon) I(e_i^T \mathbf{y} \geq 1) \geq \frac{1}{d+1}.$$

On the other hand,

$$\begin{aligned}
\sum_{i=1}^d w_i(y, \varepsilon) \sigma_i^2 g''(\theta e_i^T \mathbf{y}) &\leq \sum_{i=1}^d w_i(y, \varepsilon) \sigma_i^2 \leq \max_{i=1}^d \sigma_i^2, \\
\sum_{i=1}^d w_i(y, \varepsilon) \sigma_i^2 g'(\theta e_i^T \mathbf{y})^2 &\leq \sum_{i=1}^d w_i(y, \varepsilon) \sigma_i^2 \leq \max_{i=1}^d \sigma_i^2.
\end{aligned}$$

We conclude that

$$\begin{aligned}
&(\boldsymbol{\mu} - \boldsymbol{\mu}^+)^T Dh(\mathbf{y}) + \frac{1}{2} Tr(\Sigma D^2 h(\mathbf{y})) + \frac{1}{2} Dh(\mathbf{y})^T \Sigma Dh(\mathbf{y}) \\
&\leq \frac{-\theta}{1+d} \delta_1 + \theta^2 \frac{1}{2} \max_{i=1}^d \sigma_i^2 + \theta^2 \frac{1}{2\varepsilon} \max_{i=1}^d \sigma_i^2 + \frac{\theta^2}{2} \max_{i=1}^d \sigma_i^2 \\
&\leq \theta \max_{i=1}^d \sigma_i^2 \cdot \left(\theta \left(\frac{2\varepsilon+1}{2\varepsilon} \right) - \frac{\delta_1}{\max_{i=1}^d \sigma_i^2 (1+d)} \right) \leq -\theta \frac{\delta_1}{2(1+d)},
\end{aligned}$$

assuming that

$$\theta \leq \frac{\varepsilon}{2\varepsilon + 1} \cdot \frac{\delta_1}{(1 + d) \max_{i=1}^d \sigma_i^2}.$$

Therefore, we conclude that the condition (23) holds for $\|\mathbf{y}\|_\infty \geq 1$. On the other hand, since $Dh(\mathbf{y})^T e_i = g'(\theta y_i) = 0$, if $y_i = 0$, we also satisfy (24). Finally, we apply Lemma 9 and conclude (18). □

Now, we prove the success probability of coupon collection is uniformly bounded from 0.

Proof of Lemma 5. For any fixed $i \in \{1, \dots, d\}$, note that the event $Y_i(t) = 0$ for some $t \leq 1$ is equivalent to $L_i(1) > 0$ and hence

$$P(Y_i(t) = 0 \text{ for some } t \leq 1 | \mathbf{Y}(0) = \mathbf{y}) = P(L_i(1) > 0 | \mathbf{Y}(0) = \mathbf{y}).$$

Let $\mathbf{Z}(t) = R^{-1}(y_0 + \mathbf{X}(t))$. Define $(\mathbf{Y}^*, \mathbf{L}^*)$ to be the solution to the following Skorokhod problem:

$$\mathbf{Y}^*(t) = \mathbf{Z}(t) + \mathbf{L}^*(t) \geq 0, \mathbf{L}^*(t) = 0.$$

In particular, the process $\mathbf{L}^*(\cdot)$ is nondecreasing and $Y_i^*(t)dL_i^*(t) = 0$ for all $t \geq 0$. Then, $\mathbf{L}^*(t)$ is the minimal process that keeps $\mathbf{Y}^*(t)$ non-negative. Note that $R^{-1}\mathbf{Y}(t) = \mathbf{Z}(t) + \mathbf{L}(t) \geq 0$, therefore

$$L_i(t) \geq L_i^*(t) \text{ and } Y_i(t) \geq Y_i^*(t).$$

As a result,

$$P(Y_i(t) = 0 \text{ for some } t \leq 1 | \mathbf{Y}(0) = \mathbf{y}) = P(L_i(1) > 0 | \mathbf{Y}(0) = \mathbf{y}) \geq P(L_i^*(1) > 0 | \mathbf{Y}(0) = \mathbf{y}).$$

By definition,

$$P(L_i^*(1) > 0 | \mathbf{Y}(0) = \mathbf{y}) \geq P(Z_i(1) < 0) = P((R^{-1}\mathbf{y} + R^{-1}\boldsymbol{\mu} + R^{-1}C\mathbf{B}(1))_i < 0).$$

Note that following Assumption A1), $\|R^{-1}\mathbf{1}\|_\infty \leq b_1$ and $\mathbf{y} \leq b_1\mathbf{1}$. Therefore, $R^{-1}\mathbf{y} \leq b_1^2\mathbf{1}$. Since

$R^{-1}\boldsymbol{\mu} \leq -\delta_0$, we have $(R^{-1}\mathbf{y} + R^{-1}\boldsymbol{\mu})_i \leq b_1^2 - \delta_0$ and hence

$$P((R^{-1}\mathbf{y} + R^{-1}\boldsymbol{\mu} + R^{-1}C\mathbf{B}(1))_i < 0) \geq P((R^{-1}C\mathbf{B}(1))_i < \delta_0 - b_1^2).$$

Since $R_{ii}^{-1} \geq 1$ and $\sigma_i^2 \geq b_0^{-1}$ according to Assumption A3), $(R^{-1}C\mathbf{B}(1))_i$ is a Gaussian r.v. with variance $\geq b_0^{-1}$. Therefore, we conclude that

$$\begin{aligned} P(Y_i(t) &= 0 \text{ for some } t \leq 1 | \mathbf{Y}(0) = \mathbf{y}) \\ &\geq P((R^{-1}C\mathbf{B}(1))_i < \delta_0 - b_1^2) \\ &\geq P\left(N(0, 1) < \sqrt{b_0}(\delta_0 - b_1^2)\right) \geq p_0. \end{aligned}$$

□

We continue with the proof of Lemma 6.

Proof of Lemma 6. Recall that we have defined a sequence of Bernoulli random variables $\zeta_k(i)$ jointly with the sequence $\{\mathbf{Y}^{i,k}\}$ as

$$\zeta_k(i) = 1(Y_i(t; \mathbf{Y}^{i,k}) = 0 \text{ for some } \tau^+(k; \mathbf{Y}^{i,k}) < t \leq \tau^+(k; \mathbf{Y}^{i,k}) + 1).$$

Let $K = \min\{k : \zeta_k(i) = 1\}$. We obtain a bound for $\eta_i^1(\mathbf{y})$:

$$\eta_i^1(\mathbf{y}) \leq \sum_{k=1}^K (\tau^+(k; \mathbf{Y}^{i,k}) + 1).$$

Note that the Skorokhod mapping is monotone with respect to the initial position, i.e.,

$$\mathbf{Y}(t; \mathbf{y}^1) \geq \mathbf{Y}(t; \mathbf{y}^2) \text{ for all } t \geq 0 \text{ if } \mathbf{y}^1 \geq \mathbf{y}^2.$$

As $\mathbf{Y}(\tau^+(k; \mathbf{Y}^{i,k}); \mathbf{Y}^{i,k}) \leq b_1 \mathbf{1}$, we have

$$\mathbf{Y}^{i,k+1} \leq \mathbf{Y}(1; b_1 \mathbf{1}), \text{ and hence } \tau^+(k+1; \mathbf{Y}^{i,k+1}) \leq \tau^+(k+1; \mathbf{Y}(1; b_1 \mathbf{1})).$$

Similarly, we have

$$P(\zeta_k(i) = 1) \geq P(Y_i(t; b_1 \mathbf{1}) = 0 \text{ for some } 0 < t \leq 1) \geq p,$$

where the last inequality follows Lemma 5. As a result, we can define a Bernoulli ψ jointly with $\mathbf{Y}(1; b_1 \mathbf{1})$, such that for all $\mathbf{y} \geq 0$

$$P(\psi = 1 | \mathbf{Y}(1; b_1 \mathbf{1}) = \mathbf{y}) \leq P(Y_i(t; b_1 \mathbf{1}) = 0 \text{ for some } 0 < t \leq 1 | \mathbf{Y}(1; b_1 \mathbf{1}) = \mathbf{y}),$$

and $P(\psi = 1) = p$.

Based on the previous comparison results, we can construct a sequence of pairs $(\psi_k(i), \tau_k(i))$ to be i.i.d. copies of $(\psi, \tau^+(\mathbf{Y}(1; b_1 \mathbf{1})))$, for $1 \leq j \leq d$ and $k \geq 1$, and define $G^i = \inf\{k : \psi_k(i) = 1\}$. Then G^i is a Geometric r.v. with probability of success equal to p , and $\eta^1(\mathbf{y})$ is stochastically dominated by

$$\tau^+(\mathbf{y}) + \sum_{i=1}^d \sum_{k=1}^{G^i} (1 + \tau_k(i)).$$

Since $(\psi_k(i), \tau_k(i))$ are i.i.d., we have that

$$\sum_{i=1}^d \sum_{k=1}^{G^i} \tau_k(i) \stackrel{D}{=} \sum_{i=1}^d \sum_{k=1}^{G^i} \tilde{\tau}_k(i),$$

where for each i , $\{\tilde{\tau}_k(i) : k \geq 1\}$ is an i.i.d. sequence following the conditional distribution of $\tau_k(i)$ conditional on that $\psi_k(i) = 0$ and is independent of G^i .

The rest of the proof is to construct the r.v. Θ_d satisfying (20) and that $\tau^+(\Theta_d \mathbf{1})$ stochastically dominates $\tilde{\tau}_k(i)$.

Recall that $\mathbf{Y}(1) \leq R^{-1} \mathbf{Y}(1) \leq R^{-1} \mathbf{Y}^+(1)$ and $\bar{\mathbf{X}}(t) \leq C \mathbf{B}(t)$ for all $t > 0$, where $\mathbf{B}(t)$ is a standard Brownian motion. By the property of the Skorokhod mapping with the identity reflection

matrix, we have

$$\begin{aligned}
Y_i^+(1) &= Y_i^+(0) + X_i(1) - \left(\inf_{0 \leq t \leq 1} (Y_i^+(0) + X_i(t)) \right) \wedge 0 \\
&\leq Y_i^+(0) + X_i(1) - \left(\inf_{0 \leq t \leq 1} X_i(1) \right) \wedge 0 \\
&= Y_i^+(0) + e_i^T C \mathbf{B}(1) - \inf_{0 \leq t \leq 1} e_i^T C \mathbf{B}(t).
\end{aligned}$$

Let us write $\mathbf{U} = C \mathbf{B}(1) - \inf_{0 \leq t \leq 1} C \mathbf{B}(t)$, so whenever $\mathbf{Y}(0) = \mathbf{y} \leq b_1 \mathbf{1}$, we have

$$\mathbf{Y}(1; \mathbf{y}) \leq_{st} (b_1 + b_1 \|\mathbf{U}\|_\infty) \mathbf{1}.$$

Now we define $\Theta_d > 0$ as

$$P(\Theta_d > t) = \min \left(1, \frac{P(b_1 + b_1 \|\mathbf{U}\|_\infty > t)}{1 - p} \right) \text{ for all } t > 0.$$

Recall that $\tau_k(i)$ is a copy of $\tau^+(\mathbf{Y}(1; b_1 \mathbf{1}))$, and $\tau^+(\mathbf{y}^1) \geq_{st} \tau^+(\mathbf{y}^2)$ whenever $\mathbf{y}^1 \geq \mathbf{y}^2$. Therefore,

$$P(\tau_k(i) > t) \leq P(\tau^+((b_1 + b_1 \|\mathbf{U}\|_\infty) \mathbf{1}) > t) \leq (1 - p) P(\tau^+(\Theta_d \mathbf{1}) > t).$$

For all $t > 0$,

$$P(\tilde{\tau}_k(i) > t) = P(\tau_k(i) > t | \psi_k(i) = 0) \leq \frac{P(\tau_k(i) > t)}{1 - p} \leq P(\tau^+(\Theta_d \mathbf{1}) > t).$$

Now we show that Θ satisfies (20). Note that

$$E \exp(\theta \|\mathbf{U}\|_\infty) = \int_0^\infty \theta \exp(\theta t) P(\|\mathbf{U}\|_\infty > t) dt + 1$$

If $t = s \log(1 + d)$, breaking the integral on $[0, 1/\log(1 + d)^{1/3}]$ and $(1/\log(1 + d)^{1/3}, \infty)$, we obtain

$$\begin{aligned}
& \int_0^\infty \theta \exp(\theta t) P(\|\mathbf{U}\|_\infty > t) dt \\
&= \theta \log(1 + d) \int_0^\infty \exp(s\theta \log(1 + d)) P(\|\mathbf{U}\|_\infty > s \log(1 + d)) ds \\
&\leq \theta \log(1 + d) \exp\left(\theta \log(1 + d)^{2/3}\right) \\
&\quad + \theta \log(1 + d) \int_{1/\log(1+d)^{1/3}}^\infty \exp(s\theta \log(1 + d)) P(\|\mathbf{U}\|_\infty > s \log(1 + d)) ds.
\end{aligned}$$

Since $U_i = e_i^T C \mathbf{B}(1) - \inf_{0 \leq t \leq 1} e_i^T C \mathbf{B}(t) = \sup_{0 \leq t \leq 1} e_i^T C(\mathbf{B}(1) - \mathbf{B}(t))$ is equal in distribution to $\sup_{0 \leq t \leq 1} e_i^T C \mathbf{B}(t)$, by the reflection principle for Brownian motions, we have

$$P(U_i > t) = 2 \int_t^\infty \frac{1}{\sqrt{2\pi}\sigma_i} \exp(-r^2/2b_1\sigma_i^2) dr \leq \frac{2\sigma_i}{t\sqrt{2\pi}} \exp(-t^2/2\sigma_i^2) \leq \frac{2\sqrt{b_0}}{t\sqrt{2\pi}} \exp(-t^2/2b_0).$$

Therefore,

$$\begin{aligned}
& \int_{1/\log(1+d)^{1/3}}^\infty \exp(s\theta \log(1 + d)) P(\|\mathbf{U}\|_\infty > s \log(1 + d)) ds \\
&\leq d \int_{1/\log(1+d)^{1/3}}^\infty \frac{2\sqrt{b_0}}{s \log(1 + d) \sqrt{2\pi}} \exp\left(-\frac{s^2 \log(1 + d)^2}{2b_0} + s\theta \log(1 + d)\right) ds \\
&\leq \frac{2\sqrt{b_0}d}{\log(1 + d)^{2/3} \sqrt{2\pi}} \exp\left(-\log(1 + d)^{4/3}/3b_0\right),
\end{aligned}$$

as $\theta = o(1)$ and hence $s\theta \log(1 + d) \leq s^2 \log(1 + d)^2 / 6b_0$ for d that is large enough. Therefore, we conclude that

$$\phi_d(\theta) \leq 1 + 2(1 - p)^{-1} \theta \log(1 + d) \exp\left(\theta \log(1 + d)^{2/3}\right) + \theta O\left(d \exp\left(-\log(1 + d)^{4/3}/3b_0\right)\right).$$

□

Proof of Lemma 7. Observe that

$$\begin{aligned}
& E \exp\left(\chi(\theta) \left(\tau^+ \left(\Lambda_d^k(j) \mathbf{1}\right) + 1\right)\right) \leq \exp(\chi(\theta)) E \exp\left(h \left(\Lambda_d^k(j) \mathbf{1}, \theta\right)\right) \\
&\leq \exp(\chi(\theta) + \varepsilon \log d) E \exp(\theta \Lambda_d) = \exp(\chi(\theta) + \varepsilon \log d) \phi_d(\theta).
\end{aligned}$$

Therefore,

$$E \exp (\chi (\theta) \xi) \leq \left(\frac{\phi_d (\theta) \exp (\chi (\theta) + \varepsilon \log (d)) p}{1 - (1 - p) \phi_d (\theta) \exp (\chi (\theta) + \varepsilon \log (d))} \right)^d. \quad (26)$$

Since $\eta^n(\mathbf{y}) \leq \tau^+(y) + \xi_1 + \dots + \xi_n$ where $\tau^+(y)$, ξ_1, \dots, ξ_n are all independent of each other, and $E \exp(\chi(\theta)\tau^+(\mathbf{y})) \leq \exp(h(\mathbf{y};\theta))$ by Lemma 4, we have

$$E \exp (\chi (\theta) \eta^n (\mathbf{y})) \leq \exp (h (\mathbf{y}; \theta)) \left(\frac{\phi_d (\theta) \exp (\chi (\theta) + \varepsilon \log (d)) p}{1 - (1 - p) \phi_d (\theta) \exp (\chi (\theta) + \varepsilon \log (d))} \right)^{nd}.$$

Since the function $f(x) \doteq xp/(1 - (1 - p)x)$ is increasing in x for $x \leq 1/(1 - p)$, under (22), we have

$$\frac{[\phi_d (\theta) \exp (\chi (\theta) + \varepsilon \log (d))] p}{1 - (1 - p) [\phi_d (\theta) \exp (\chi (\theta) + \varepsilon \log (d))]} \leq f \left(\frac{1}{(1 - p)(1 + p)} \right) = \frac{1}{1 - p},$$

and we are done. □

We conclude this section with the proof of Lemma 8.

Proof of Lemma 8. Let us write $\xi_0 = \tau_b^+(\mathbf{y})$, $A_{-1} = 0$, and $A_n = \xi_0 + \dots + \xi_n$

$$\bar{N}(t) = \sup\{n \geq -1 : A_n \leq t\},$$

so that $\bar{N}(\cdot)$ is a delayed renewal process. Following Lemma 6, we have $\bar{N}(t) \leq_{st} \mathcal{N}(t; \mathbf{y})$ and, therefore, for any $\beta > 0$,

$$E \exp (-\beta \mathcal{N}(t; \mathbf{y})) \leq E \exp (-\beta \bar{N}(t)).$$

According to Lemma 6 and Lemma 7,

$$M_n = \exp (\chi (\theta) A_n - h (\mathbf{y}; \theta)) (1 - p)^{dn}$$

is a non-negative supermartingale and, therefore,

$$1 \geq EM_{\bar{N}(t)+1} \geq E \left(\exp (\chi (\theta) t - h (\mathbf{y}; \theta)) (1 - p)^{(\bar{N}(t)+1)d} \right),$$

thereby concluding that

$$E \left((1-p)^{d\mathcal{N}(t;\mathbf{y})} \right) \leq E \left((1-p)^{d\bar{N}(t)} \right) \leq \exp(h(\mathbf{y};\theta)) \cdot \exp(-\chi(\theta)t) \cdot (1-p)^{-d},$$

and the result follows. \square

5 Step 3: Concluding the Proof of Theorem 1

For any $f \in \mathcal{L}$,

$$\begin{aligned} & E |f(\mathbf{Y}(t;\mathbf{y})) - f(\mathbf{Y}(t;\mathbf{Y}(\infty)))| \\ & \leq E \|\mathbf{Y}(t;\mathbf{y}) - \mathbf{Y}(t;\mathbf{Y}(\infty))\|_\infty \\ & \leq E \|\mathbf{Y}(t;\mathbf{y}) - \mathbf{Y}(t;\mathbf{0})\|_1 + E \|\mathbf{Y}(t;\mathbf{0}) - \mathbf{Y}(t;\mathbf{Y}(\infty))\|_1. \end{aligned}$$

Therefore, by Lemma 3, we have that

$$\begin{aligned} & E |f(\mathbf{Y}(t;\mathbf{y})) - f(\mathbf{Y}(t;\mathbf{Y}(\infty)))| \\ & \leq d \cdot \kappa_0 \cdot \left(E \left((1-\beta_0)^{\mathcal{N}(t;\mathbf{y})} \|\mathbf{y}\|_1 \right) + E \left((1-\beta_0)^{\mathcal{N}(t;\mathbf{Y}(\infty))} \|\mathbf{Y}(\infty)\|_1 \right) \right). \end{aligned} \tag{27}$$

For the last term, according to the Cauchy-Schwarz inequality, we have that

$$E \left((1-\beta_0)^{\mathcal{N}(t;\mathbf{Y}(\infty))} \|\mathbf{Y}(\infty)\|_1 \right) \leq E^{1/2} \left(\|\mathbf{Y}(\infty)\|_1^2 \right) E^{1/2} \left((1-\beta_0)^{2\mathcal{N}(t;\mathbf{Y}(\infty))} \right).$$

Following the stochastic domination result (16) and the fact that $R^{-1} \geq I$, we have

$$\|\mathbf{Y}(\infty)\|_1 \leq \|R^{-1}\mathbf{Y}(\infty)\|_1 \leq \|R^{-1}\mathbf{Y}^+(\infty)\|_1 \leq \|R^{-1}\|_1 \|\mathbf{Y}^+(\infty)\|_1 \leq b_1 \|\mathbf{Y}^+(\infty)\|_1.$$

Moreover,

$$\|\mathbf{Y}^+(\infty)\|_1^2 = \left(\sum_{i=1}^d Y_i^+(\infty) \right)^2 \leq d \left(\sum_{i=1}^d Y_i^+(\infty)^2 \right).$$

By definition, $Y_i^+(\infty)$ represents a one-dimensional RBM, with dirft $-(\mu_i^+ - \mu_i)$ and variance σ_i^2 , in its steady state. So $Y_i^+(\infty)$ follows an exponential distribution with mean $\sigma_i^2/2 (\mu_i^+ - \mu_i)$, and therefore (recall that we have chosen $\delta_1 = \delta_0\beta_0/2\kappa_0$),

$$E\left(Y_i^+(\infty)^2\right) = \frac{\sigma_i^2}{(\mu_i^+ - \mu_i)} = \frac{\sigma_i^2}{\delta_1} \leq \frac{2\sigma_i^2\kappa_0}{\delta_0\beta_0} \leq \frac{2b_0\kappa_0}{\delta_0\beta_0},$$

which concludes that

$$E^{1/2}\left(\|\mathbf{Y}(\infty)\|_1^2\right) \leq \sqrt{2} \cdot d \cdot \frac{\kappa_0^{1/2}}{\delta_0^{1/2}\beta_0^{1/2}} b_0^{1/2}. \quad (28)$$

Next, invoking Proposition 1 with $\beta \in (0, \min(\beta_0, 1/3) \cdot 1/3)$, we can guarantee that $(1 - \beta) \geq (1 - \beta_0)^2$, and therefore conclude that

$$\begin{aligned} E\left((1 - \beta_0)^{2\mathcal{N}(t; \mathbf{Y}(\infty))}\right) &\leq E\left[\exp(\zeta_0 \|\mathbf{Y}(\infty)\|_\infty / (d^3 \log(d)) + \beta/d^2)\right] \times \\ &\quad \exp(-\zeta_1 t / (d^4 \log(d))) \cdot (1 - \beta)^{-1}, \end{aligned} \quad (29)$$

where

$$\zeta_0 = \frac{\delta_1 \cdot \beta}{\max_{i=1}^d \sigma_i^2}, \quad \zeta_1 = \frac{\delta_1^2 \cdot \beta}{\max_{i=1}^d \sigma_i^2}.$$

Once again, using the stochastic domination result (16), we have that

$$\|\mathbf{Y}(\infty)\|_\infty \leq \|R^{-1}\mathbf{Y}(\infty)\|_\infty \leq \|R^{-1}\mathbf{Y}^+(\infty)\|_\infty \leq b_1 \|\mathbf{Y}^+(\infty)\|_\infty.$$

Observe that

$$P(\|\mathbf{Y}^+(\infty)\|_\infty > t) \leq \sum_{i=1}^d P(Y_i^+(\infty)_\infty > t) \leq d \exp\left(-\frac{2\delta_1}{\max_{i=1}^d \sigma_i^2} t\right).$$

We conclude that

$$\begin{aligned} &E[\exp(\zeta_0 \|\mathbf{Y}(\infty)\|_\infty / (d^3 \log(d)))] \\ &\leq \frac{\zeta_0}{d^3 \log(d)} \int_0^\infty \exp\left(\frac{\zeta_0}{(d^3 \log(d))} t\right) P(\|\mathbf{Y}^+(\infty)\|_\infty > t) dt + 1 \\ &\leq \frac{\zeta_0 d}{d^3 \log(d)} \int_0^\infty \exp\left(\frac{\zeta_0}{(d^3 \log(d))} t - \frac{2\delta_1}{\max_{i=1}^d \sigma_i^2 b_1} t\right) dt + 1. \end{aligned}$$

Hence, using this estimate, together with (28) and (29) we conclude that

$$E^{1/2} \left(\|\mathbf{Y}(\infty)\|_1^2 \right) E^{1/2} \left((1 - \beta_0)^{2\mathcal{N}(t; \mathbf{Y}(\infty))} \right) \leq 3 \cdot d \cdot \frac{\kappa_0^{1/2}}{\delta_0^{1/2} \beta_0^{1/2}} b_0^{1/2} \exp \left(-\frac{\zeta_1}{2(d^4 \log(d))} t \right).$$

On the other hand, directly from Proposition 1, we obtain (with the same selection of β , in particular $\beta \in (0, 1/3)$) that

$$\|\mathbf{y}\|_1 E \left((1 - \beta_0)^{\mathcal{N}(t; \mathbf{y})} \right) \leq 3 \cdot \|\mathbf{y}\|_1 \exp \left(\zeta_0 \|\mathbf{y}\|_\infty / (d^3 \log(d)) \right) \cdot \exp \left(-\zeta_1 t / (d^4 \log(d)) \right).$$

Putting these estimates together in (27), we obtain that

$$\begin{aligned} & E |f(\mathbf{Y}(t; \mathbf{y})) - f(\mathbf{Y}(t; \mathbf{Y}(\infty)))| \\ & \leq 3 \cdot d \cdot \exp \left(-\frac{\zeta_1}{d^4 \log(d)} t \right) \left(\|\mathbf{y}\|_1 \cdot \kappa_0 \cdot \exp \left(\zeta_0 \frac{\|\mathbf{y}\|_\infty}{d^3 \log(d)} \right) + \frac{\kappa_0^{1/2}}{\delta_0^{1/2} \beta_0^{1/2}} b_0^{1/2} \right). \end{aligned}$$

Acknowledgement: Blanchet acknowledges support from the NSF through the grant DMS-132055.

References

- [1] A. Budhiraja and C. Lee. Stationary distribution convergence for generalized Jackson networks in heavy traffic. *Mathematics of Operations Research*, 34:45–56, 2009.
- [2] H. Chen and D. D. Yao. *Fundamentals of Queueing Networks: Performance, Asymptotics and Optimization*. Springer Verlag, 2001.
- [3] D. Gamarnik and A. Zeevi. Validity of heavy traffic steady-state approximations in generalized Jackson networks. *Annals of Applied Probability*, 16:56–90, 2006.
- [4] J. M. Harrison and M. I. Reiman. Reflected Brownian motion on an orthant. *Annals of Applied Probability*, 9:302–308, 1981.
- [5] J. M. Harrison and M. I. Reiman. Brownian models of open queueing networks with homogeneous customer populations. *Stochastics*, 2:77–115, 1987.

- [6] O. Kella. Stability and nonproduct form of stochastic fluid networks with Lévy inputs. *Annals of Applied Probability*, 6:186–199, 1996.
- [7] O. Kella and S. Ramasubramanian. Asymptotic irrelevance of initial conditions for Skorokhod refection mapping on the nonnegative orthant. *Mathematics of Operations Research*, 37:301–312, 2012.
- [8] O. Kella and W. Whitt. Stability and structural properties of stochastic storage networks. *Journal of Applied Probability*, 33:1169–1180, 1996.